



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

PATENT

Applicant : Kenji Nakamura et al.
Serial No.: 10/620,499
Filed: July 16, 2003
Title: "VOICE RECOGNITION APPARATUS"
Docket No.: 35907

LETTER

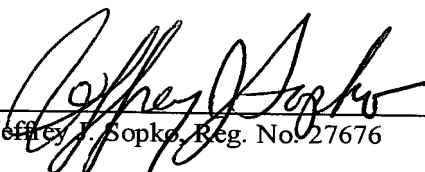
Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir/Madam:

Enclosed is a certified copy of Japan Patent Application No. 2002-208386, filed July 17, 2002; the priority of which has been claimed in the above-identified application.

Respectfully submitted,

PEARNE & GORDON LLP


Jeffrey J. Sopko, Reg. No. 27676

526 Superior Avenue East
Suite 1200
Cleveland, Ohio 44114-1484
(216) 579-1700

September 12, 2003

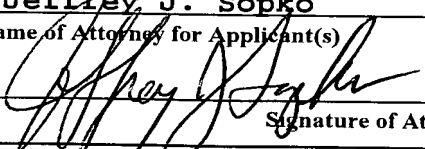
I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner of Patents, P.O. Box 1450, Alexandria, Va. 22313-1450 on the date indicated below.

Jeffrey J. Sopko

Name of Attorney for Applicant(s)

09/12/2003

Date


Signature of Attorney

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 2 年 7 月 1 7 日
Date of Application:

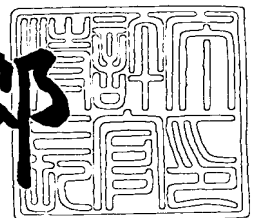
出 願 番 号 特 願 2 0 0 2 - 2 0 8 3 8 6
Application Number:
[ST. 10/C]: [J P 2 0 0 2 - 2 0 8 3 8 6]

出 願 人 松 下 電 器 産 業 株 式 有 限 公 司
Applicant(s):

2 0 0 3 年 7 月 8 日

特許庁長官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特 2 0 0 3 - 3 0 5 3 7 0 5

【書類名】 特許願

【整理番号】 2913040379

【提出日】 平成14年 7月17日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 9/00

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 中村 賢二

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 原田 博司

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 緒方 芳幸

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 立山 雅一

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 五島 龍宏

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 西岡 靖幸

【発明者】

【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

【氏名】 黒木 義明

【特許出願人】

【識別番号】 000005821

【氏名又は名称】 松下電器産業株式会社

【代理人】

【識別番号】 100097445

【弁理士】

【氏名又は名称】 岩橋 文雄

【選任した代理人】

【識別番号】 100103355

【弁理士】

【氏名又は名称】 坂口 智康

【選任した代理人】

【識別番号】 100109667

【弁理士】

【氏名又は名称】 内藤 浩樹

【手数料の表示】

【予納台帳番号】 011305

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9809938

【書類名】 明細書

【発明の名称】 音声認識装置

【特許請求の範囲】

【請求項 1】 キーボードなどの入力装置と、認識結果を表示する表示装置と、前記入力装置と前記表示装置が接続され、音声認識装置全体の制御を行う信号処理部と、前記信号処理部より入力された音声信号を分割した音響データと内部で保持している複数の単語音響データとを比較し、前記分割した音響データに一致する単語音響データの単語識別子を含む識別情報を音声認識の結果として前記信号処理部へ出力する音声認識処理部とを有する不特定話者方式の音声認識装置であって、

前記音声認識処理部は、既に登録されている音素モデルから生成される単語音響データと話者が発した名前から生成される音響データとを逐次比較し、前記単語音響データに合致する前記音響データが生成されたとき、前記生成された音響データに対応する音響データ識別子をトレーニング信号として蓄積することを特徴とする音声認識装置。

【請求項 2】 前記信号処理部は、名前が発音される前後の言葉を記憶するメモリ部を有し、名前の前の言葉を開始信号、後の言葉を終了信号とすることを特徴とする請求項 1 に記載の音声認識装置。

【請求項 3】 前記信号処理部は、日本語において名前の前に必ず存在する無音を記憶するメモリ部を有し、前記無音を検知して開始信号とすることを特徴とする請求項 2 に記載の音声認識装置。

【請求項 4】 前記信号処理部は、無音を検知する検知部と、前記検知した無音を開始信号とする制御部とを有することを特徴とする請求項 3 に記載の音声認識装置。

【請求項 5】 前記信号処理部は、無音検知のためのスレッシュホールドレベルを前記検知部に設けたことを特徴とする請求項 4 に記載の音声認識装置。

【請求項 6】 前記音声認識処理部は、単語を単語辞書格納部に格納する際、単語の先頭部分とその他の部分に分け、前記先頭部分が同じ物をグループ化して格納することを特徴とする請求項 1 に記載の音声認識装置。

【請求項 7】前記音声認識処理部は、あらかじめ前記単語辞書格納部内の先頭文字と音素モデルから先頭文字の単語音響データを作成し、言語モデル作成格納部に格納することを特徴とする請求項 6 に記載の音声認識装置。

【請求項 8】前記音声認識処理部は、単語辞書の構成を先頭文字と姓と名前とにブロック分けすることを特徴とする請求項 6 に記載の音声認識装置。

【請求項 9】前記音声認識処理部は、言語モデル作成格納部に格納された各単語音響データに対して、使用されたとき単語識別部より出力される一致度で一番高いものに頻度 1 を与え、使用のたびに加算して重み付けを行うことを特徴とする請求項 1 に記載の音声認識装置。

【請求項 10】前記音声認識処理部は、或る任意の頻度以上の単語音響データのみを使って認識動作を行わせることを特徴とする請求項 9 に記載の音声認識装置。

【請求項 11】前記音声認識処理部は、使用頻度の高い順に任意の単語数ごとにブロック化し、使用頻度の高いブロックの単語音響データを出力させ、ブロック毎に表示させることを特徴とする請求項 9 に記載の音声認識装置。

【請求項 12】前記信号処理部は時計機能を有し、前記音声認識処理部は前記信号処理部から通知される時刻に基づいて使用頻度を算出する期限を設けることを特徴とする請求項 9 乃至 11 のいずれか 1 に記載の音声認識装置。

【請求項 13】前記信号処理部は、認識動作後に表示装置に表示された結果がユーザの意図する結果と異なった場合には異なったことを示す情報を内蔵するメモリ部に蓄積し、次ぎに同じ発音があった場合に、前記異なったことを示す情報をもとに、一度間違った情報は二度と前記表示装置に表示しないことを特徴とする請求項 1 に記載の音声認識装置。

【請求項 14】前記信号処理部は、認識動作後に表示装置に表示された結果がユーザの意図する結果と異なった場合には異なったことを示す情報を前記音声認識処理部のメモリ部に蓄積し、次ぎに同じ発音があった場合に、前記異なったことを示す情報をもとに、一度間違った情報は二度と前記表示装置に表示しないことを特徴とする請求項 1 に記載の音声認識装置。

【請求項 15】前記音声認識処理部は、仕向け地ごとの或る特定の言葉を単語

辞書格納部にあらかじめ持たせることを特徴とする請求項 1 に記載の音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、不特定話者の音声を認識する音声認識装置に関するものである。

【0002】

【従来の技術】

近年、電話機やファクシミリ装置、カーナビゲーション装置などの情報処理装置であって、音声入力による本体操作が可能なものが製品化されるようになってきた。いわゆる音声認識技術を応用した製品群である。音声認識技術の方式は、話者を限定しない不特定話者方式 (speaker independent) と、話者を限定する特定話者方式 (speaker dependent) との二つに大別される。不特定話者方式は、音声に含まれる言語的な特徴を抽出し、ニューラルネットワークに代表されるパターン認識技術を応用して話者の発話内容を推定するものである。ところが、話者の発話音声には各個人特有の声質があり、不特定の話者に対して安定した認識率や認識スピードを確保するためには、CPU の高性能化やメモリの大容量化を図る必要があり、結果として製品のコストアップにつながる。一方、特定話者方式は装置の初回使用時に話者自身の声質を登録 (トレーニング) することが必要であり、不特定話者方式と比べると、話者への利便性に欠けると言う問題はあるが、不特定話者方式よりは安価で、認識率が高く、認識スピードが速い装置を構築することができる。このように両方式とも、一長一短の性能を持っており、また、認識する単語数が多くなれば成る程、CPU の高性能化やメモリの大容量化の必要性が生じてくる。音声認識処理では、あらかじめデータベースの形で音声認識装置内に保存された単語群の中から、話者が発声した単語に該当するものを識別し、結果を話者に返すことが基本的な動作となる。以下、図面を参照しながら従来の特定話者方式での音声認識装置についておおまかな動作説明を行う。

【0003】

図 9 は特定話者方式を用いる従来の音声認識装置を示すブロック図であり、図 10 は図 9 の音声認識処理部を示すブロック図、図 11 は図 10 の単語音響データ格納部を示すブロック図である。

【0004】

図 9 ～図 11 において、1 はマイク（マイクロフォン）、2 はスピーカ、3 は各種データを入力するための入力装置、4 は文字や図形を表示する表示装置、5 は入力装置 3 および表示装置 4 が接続され、音声認識装置全体の制御を行う信号処理部、6 は信号処理部 5 との間で情報のやり取りを行う音声認識処理部、7 は信号処理部 5 からの音声信号を入力して音響データを出力する音響処理部、8 は信号処理部 5 からの単語識別子と音響処理部 7 からの音響データを入力して単語識別子と単語音響データを出力する単語音響データ格納部、9 は単語音響データ格納部 8 からの単語識別子と単語音響データを入力して選択された単語識別子を識別情報として出力する単語識別部である。

【0005】

このように構成された音声認識装置について、その動作を説明する。

【0006】

話者の発声した単語は、マイク 1 で電気信号へ変換され、信号処理部 5 へ入力される。信号処理部 5 では、入力された音声信号を音声認識処理部 6 での処理に適した形式の音声信号へ変換する。音声認識処理部 6 において、音響処理部 7 は、信号処理部 5 が出力する音声信号から音響的な特徴量を抽出し、音響データとして単語識別部 9 へと出力する。単語識別部 9 では、入力された音響データにもっとも一致するものを単語音響データ格納部 8 にあらかじめ保持されている音響データの中から探し出す。この結果一致した音響データに関連づけられた単語識別子が識別情報として信号処理部 5 へと戻される。信号処理部 5 では、音声認識の結果である識別情報によって話者の発声した単語を認識でき、それに基づいて装置の適切な処理制御を実施したり、表示装置 4 を介して話者に認識結果をフィードバックする。入力装置 3 は、話者が認識結果の確認を行うためのキー入力や装置全体の制御を行うための一般的な入力装置である。

【0007】

前述したように話者特定方式においては単語音響データはトレーニングによって生成される。したがって、装置の初期状態では単語音響データは未定義の状態であるため、音声認識処理の前にこのトレーニングが必須となる。トレーニングとは、話者が認識対象であるすべての単語について発声を行い、それを単語音響データ格納部 8 に登録する処理である。トレーニングにおいて、話者が発声した特定の認識対象の単語は、マイク 1 により入力され、信号処理部 5 によって音声信号に変換されるが、このとき個々の認識対象単語を区別するための単語識別子が付加される。信号処理部 5 からの音声信号は、音響処理部 7 で音響データに変換され、単語音響データ 11 として単語識別子 10 とともに単語音響データ格納部 8 へ供給される。単語音響データ格納部 8 ではこの単語音響データ 11 と単語識別子 10 が互いに関連付けて格納される。こうして全ての音声認識対象の単語に対して同様のトレーニングを繰り返すことにより初めて音声認識が可能になる。

【0008】

図 12 は不特定話者方式を用いる従来の音声認識装置を示すブロック図であり、図 13 は図 12 の音声認識処理部 6 を示すブロック図、図 14 は図 13 の単語辞書格納部を示すブロック図である。

【0009】

図 12、図 13 において、マイク 1、スピーカ 2、入力装置 3、表示装置 4、信号処理部 5、音声認識処理部 6、音響処理部 7、単語識別部 9 は図 9、図 10 と同様のものなので、同一符号を付し、説明は省略する。12 は単語識別子と単語データを入力し単語辞書として格納する単語辞書格納部、13 は音素データを出力する音素モデル格納部、14 は単語辞書格納部 12 からの単語識別子および単語データと音素モデル格納部 13 からの音素データとを入力して単語識別子と単語音響データとから成る言語モデルを作成し格納する言語モデル作成格納部である。

【0010】

このように構成された音声認識装置について、その動作を説明する。

【0011】

不特定話者方式での音声認識では、単語辞書格納部 12 には何も格納されておらず、話者は、使用に当り入力装置 3 を使い単語データを入力する必要がある。入力された単語データは、信号処理部 5 に入力され、信号処理部 5 で単語識別子を付加され、音声認識処理部 6 の単語辞書格納部 12 に入力され保持される。話者の発声した単語は、マイク 1 で電気信号へ変換され、信号処理部 5 へ入力される。信号処理部 5 では、入力された音声信号を音声認識処理部 6 での処理に適した形式の音声信号へ変換する。音声認識処理部 6 の音響処理部 7 は、この音声信号から音響的な特徴量を抽出し音響データとして単語識別部 9 へと出力する。一方、音素モデル格納部 13 には、通常使用される国の言語に合わせた音素モデルが音素データとして格納されており、認識動作が開始されると同時に言語モデル作成格納部 14 に音素データが入力される。

【0012】

言語モデル作成格納部 14 では、入力された単語データと音素データで単語音響データを生成し、単語識別子と共に単語識別部 9 へ出力する。また、これを単語辞書格納部 12 に格納された単語データ全てにおいて繰り返す。単語識別部 9 では、入力された音響データにもっとも一致するものを言語モデル作成格納部 14 で順次作成される単語音響データの中から探し出す。この結果一致した単語音響データに関連づけられた単語識別子が識別情報として信号処理部 5 へと戻される。信号処理部 5 では、音声認識の結果である識別情報によって話者の発声した単語を認識でき、それに基づいて装置の適切な処理制御を実施したり、表示装置 4 を介して話者に認識結果をフィードバックする。入力装置 3 は、話者が認識結果の確認を行うためのキー入力や装置全体の制御を行うための一般的な入力装置である。

【0013】

このように、不特定話者方式での音声認識装置では、単語辞書それぞれについて音素モデルより単語音響データを生成するため、処理速度の高速化やメモリの大容量化が求められ、また、特定話者方式での音声認識に比べ、認識率、認識スピードが劣るといった不具合がある。

【0014】

【発明が解決しようとする課題】

このように、従来の音声認識装置においては、特定話者方式では、認識率が高く認識スピードが速いと言う利点はあるものの、トレーニング作業を行う必要があり、話者に負担がかかるという問題点を有していた。一方、不特定話者方式では、トレーニング作業が不要と言う利点はあるものの、認識率が低く認識スピードが遅いという問題点を有していた。このように、両方式とも一長一短あり、利便性に欠ける等の問題点を有していた。

【0015】

この音声認識装置では、電話による会話の最初に相手の名前が発せられる頻度が高いことを利用して、話者に意識させることなくトレーニングを行うことができ、使用して行くにつれて特定話者方式の認識率が向上し、認識スピードが高速化することが要求されている。

【0016】

本発明は、この要求を満たすため、話者に意識させることなくトレーニングを行うことができ、使用して行くにつれて特定話者方式の認識率と認識スピードを高めることができる音声認識装置を提供することを目的とする。

【0017】**【課題を解決するための手段】**

上記課題を解決するために本発明の音声認識装置は、キーボードなどの入力装置と、認識結果を表示する表示装置と、入力装置と表示装置が接続され、音声認識装置全体の制御を行う信号処理部と、信号処理部より入力された音声信号を分割した音響データと内部で保持している複数の単語音響データとを比較し、分割した音響データに一致する単語音響データの単語識別子を含む識別情報を音声認識の結果として信号処理部へ出力する音声認識処理部とを有する不特定話者方式の音声認識装置であって、音声認識処理部は、既に登録されている音素モデルから生成される単語音響データと話者が発した名前から生成される音響データとを逐次比較し、単語音響データに合致する音響データが生成されたとき、生成された音響データに対応する音響データ識別子をトレーニング信号として蓄積する構成を備えている。

【0018】

これにより、話者に意識させることなくトレーニングを行うことができ、使用して行くにつれて特定話者方式の認識率と認識スピードを高めることができる音声認識装置が得られる。

【0019】

【発明の実施の形態】

本発明の請求項1に記載の音声認識装置は、キーボードなどの入力装置と、認識結果を表示する表示装置と、入力装置と表示装置が接続され、音声認識装置全体の制御を行う信号処理部と、信号処理部より入力された音声信号を分割した音響データと内部で保持している複数の単語音響データとを比較し、分割した音響データに一致する単語音響データの単語識別子をを含む識別情報音声認識の結果として信号処理部へ出力する音声認識処理部とを有する不特定話者方式の音声認識装置であって、音声認識処理部は、既に登録されている音素モデルから生成される単語音響データと話者が発した名前から生成される音響データとを逐次比較し、単語音響データに合致する音響データが生成されたとき、生成された音響データに対応する音響データ識別子をトレーニング信号として蓄積することとしたものである。

【0020】

この構成により、話者に意識させずにトレーニングを行うことができるので、話者の負担になるトレーニングを排除することができ、また、不特定話者方式でありながら、特定話者方式の特徴である高い認識率と認識スピードを得ることができるという作用を有する。

【0021】

請求項2に記載の音声認識装置は、請求項1に記載の音声認識装置において、信号処理部は、名前が発音される前後の言葉を記憶するメモリ部を有し、名前の前の言葉を開始信号、後の言葉を終了信号とすることとしたものである。

【0022】

この構成により、名前が発音される前後の言葉の一致度が向上し、トレーニング精度を更に向上させることができるという作用を有する。

【0023】

請求項3に記載の音声認識装置は、請求項2に記載の音声認識装置において、信号処理部は、日本語において名前の前に必ず存在する無音を記憶するメモリ部を有し、無音を検知して開始信号とすることとしたものである。

【0024】

この構成により、無音を名前の開始信号とすることができるので、メモリ部における必要メモリ量を低減することができると共に、名前のトレーニング精度を更に向上させることができるという作用を有する。

【0025】

請求項4に記載の音声認識装置は、請求項3に記載の音声認識装置において、信号処理部は、無音を検知する検知部と、検知した無音を開始信号とする制御部とを有することとしたものである。

【0026】

この構成により、無音検知を信号処理部で行うので、音声認識処理部の負荷を軽減することができ、音声認識処理部におけるトレーニング処理をより向上させることができるという作用を有する。

【0027】

請求項5に記載の音声認識装置は、請求項4に記載の音声認識装置において、信号処理部は、無音検知のためのスレッショールドレベルを検知部に設けることとしたものである。

【0028】

この構成により、雑音による誤動作発生を防ぐことができるという作用を有する。

【0029】

請求項6に記載の音声認識装置は、請求項1に記載の音声認識装置において、音声認識処理部は、単語を単語辞書格納部に格納する際、単語の先頭部分とその他の部分に分け、先頭部分が同じ物をグループ化して格納することとしたものである。

【0030】

この構成により、先頭部分が同じグループのみを言語モデル化して識別を行うようにすることができるので、識別動作の高速化を図ることができ、単語辞書格納部のメモリ量の削減を図ることができるという作用を有する。

【0031】

請求項7に記載の音声認識装置は、請求項6に記載の音声認識装置において、音声認識処理部は、あらかじめ単語辞書格納部内の先頭文字と音素モデルから先頭文字の単語音響データを作成し、言語モデル作成格納部に格納することとしたものである。

【0032】

この構成により、単語辞書格納部からの単語データの呼び出しや言語モデル作成格納部からの音素データの呼び出し、またそれらによる単語音響データの作成時間を省くことができ、更なる高速化が図れるという作用を有する。

【0033】

請求項8に記載の音声認識装置は、請求項6に記載の音声認識装置において、音声認識処理部は、単語辞書の構成を先頭文字と姓と名前とにブロック分けすることとしたものである。

【0034】

この構成により、ブロック分けが更に進められ、使用メモリ量をさらに削減することができるという作用を有する。

【0035】

請求項9に記載の音声認識装置は、請求項1に記載の音声認識装置において、音声認識処理部は、言語モデル作成格納部に格納された各単語音響データに対して、使用されたとき単語識別部より出力される一致度で一番高いものに頻度1を与え、使用のたびに加算して重み付けを行うこととしたものである。

【0036】

この構成により、重み付けの高い順に単語音響データを並べ替えることができるので、音声認識時の高速化を図ることができるという作用を有する。

【0037】

請求項10に記載の音声認識装置は、請求項9に記載の音声認識装置において

、音声認識処理部は、或る任意の頻度以上の単語音響データのみを使って認識動作を行わせることとしたものである。

【 0 0 3 8 】

この構成により、或る任意の値以下の頻度を持つ単語音響データは言語モデル作成格納部に出力しないので、音声認識における高速化を図ることができるという作用を有する。

【 0 0 3 9 】

請求項 1 1 に記載の音声認識装置は、請求項 9 に記載の音声認識装置において、音声認識処理部は、使用頻度の高い順に任意の単語数ごとにブロック化し、使用頻度の高いブロックの単語音響データを出力させ、ブロック毎に表示させることとしたものである。

【 0 0 4 0 】

この構成により、使用頻度が低い音声が入力されても認識が可能で、またブロック毎の識別動作で表示する事により高速な認識動作を行うことができるという作用を有する。

【 0 0 4 1 】

請求項 1 2 に記載の音声認識装置は、請求項 9 乃至 1 1 のいずれか 1 に記載の音声認識装置において、信号処理部は時計機能を有し、音声認識処理部は信号処理部から通知される時刻に基づいて使用頻度を算出する期限を設けることとしたものである。

【 0 0 4 2 】

この構成により、使用頻度が最も高いものであっても期限外ならば低位に位置することができるので、ユーザにとって近似点で最も使用頻度が高いものから順に認識動作を行うようにすることができ、より高速化を図ることができるという作用を有する。

【 0 0 4 3 】

請求項 1 3 に記載の音声認識装置は、請求項 1 に記載の音声認識装置において、信号処理部は、認識動作後に表示装置に表示された結果がユーザの意図する結果と異なった場合には異なったことを示す情報を内蔵するメモリ部に蓄積し、次

ぎに同じ発音があった場合に、異なったことを示す情報をもとに、一度間違った情報は二度と表示装置に表示しないこととしたものである。

【0044】

この構成により、一度間違った情報は二度と表示装置に表示しないので、ユーザにとっての利便性が向上するという作用を有する。

【0045】

請求項14に記載の音声認識装置は、請求項1に記載の音声認識装置において、信号処理部は、認識動作後に表示装置に表示された結果がユーザの意図する結果と異なった場合には異なったことを示す情報を音声認識処理部のメモリ部に蓄積し、次ぎに同じ発音があった場合に、異なったことを示す情報をもとに、一度間違った情報は二度と表示装置に表示しないこととしたものである。

【0046】

この構成により、請求項13と同様の作用を有すると共に、音声認識処理部のメモリ部に情報を蓄積するので、信号処理部の負荷およびメモリ量を減らすことができるという作用を有する。

【0047】

請求項15に記載の音声認識装置は、請求項1に記載の音声認識装置において、音声認識処理部は、仕向け地ごとの或る特定の言葉を単語辞書格納部にあらかじめ持たせることとしたものである。

【0048】

この構成により、ユーザが装置を使用する最初に或る特定の言葉を発音して、自動的に言語を選択することができるという作用を有する。

【0049】

以下、本発明の実施の形態について、図1～図8を参照しながら説明する。

【0050】

(実施の形態1)

図1は、本発明の実施の形態1による音声認識装置を示すブロック図であり、不特定話者方式での音声認識装置を示す。

【0051】

図1において、マイク1、スピーカ2、入力装置3、表示装置4、信号処理部5、音声認識処理部6、音響処理部7、単語識別部9、単語辞書格納部12、音素モデル格納部13、言語モデル作成格納部14は図12、図13と同様のものなので、同一符号を付し、説明は省略する。16は音響データ識別子と音響データを格納するメモリ部である。

【0052】

このように構成された音声認識装置について、電話機を例に取り、話者が意識しない自動トレーニングについて説明する。

【0053】

通常、話者が他者に電話をかけた際、会話の最初に相手の名前が発せられる頻度は非常に高い傾向にある。例として、日本語では『もしもし中村ですが、松下さんを、お願いします。』、また英語では『Hello. This is Nakamura. Mr. Matsushita Please.』などである。上記日本語の例を取り音声認識部の動作を説明すると、図1において『もしもし中村ですが、松下さんを、お願いします。』なる音声信号がマイク1より信号処理部5に入力される。この音声信号を入力した音響処理部7では、『もしもし中村ですが、松下さんを、お願いします。』の音声のある任意の時間間隔で『もし』『もし』『なか』『むら』『です』『が、』『まつ』『した』『さん』『を、』『おね』『がい』『しま』『す。』と分割し、分割した音響データ（単語音響データ）としてメモリ部16に出力する。また、その分割した音響データそれぞれに対して、音響データ識別子が信号処理部5より割り振られる。メモリ部16は、信号処理部5より入力された音響データ識別子に音響処理部7で生成された音響データを対応させ蓄積する。次に、メモリ部16は、この蓄積された音響データとそれに対応する音響データ識別子を単語認識部9に出力する。

【0054】

一方、単語辞書格納部12では、現在電話をしている相手に相当する単語データ『まつした』が単語識別子よりすでに判明しているため、単語識別子と単語データを言語モデル作成格納部14に出力する。これと同時に音素モデル格納部13より音素データが言語モデル作成格納部14に出力され、言語モデル作成格納

部 14 で単語音響データが作成され、単語識別子と共に、単語識別部 9 に出力される。単語識別部 9 では、言語モデル作成格納部 14 から出力される単語音響データ『まつした』とメモリ部 16 から出力される音響データ『もし』『もし』『なか』『むら』『です』『が、』『まつ』『した』『さん』『を、』『おね』『がい』『しま』『す。』を比較識別し、一致度が高い『まつ』『した』の音響データ識別子を識別情報として信号処理部 5 に出力する。信号処理部 5 は、一致度が高い『まつ』『した』の音響データ識別子と制御信号をメモリ部 16 に出力し、メモリ部 16 はこれを受け、言語モデル作成格納部 14 に、音響データ識別子とこれに対応する音響データを出力する。言語モデル作成格納部 14 では、入力された音響データ識別子を任意の識別子に置き換え、音響データは時間的に連続になるよう結合され格納保持する。このようにして次に話者が『松下』と発音した場合は、言語モデル作成格納部 14 は、格納保持した単語音響データと単語識別子をまず最初に単語識別部 9 へ出力して認識動作を行わせ、或る任意の一致度以上の場合、単語識別部 9 は、単語識別子を含む識別情報を信号処理部 5 へ出力し、信号処理部 5 は、表示装置 4 に情報を出力する。或る任意の一致度以下の場合、従来の音素モデルをもとに単語音響データを作成すると言う複雑な処理を行う。

【0055】

以上により、不特定話者方式の音声認識装置でありながら、使用するにつれて認識率や認識スピードが向上して行き、話者にとって非常に利便性の優れた音声認識装置を得ることができる。

【0056】

(実施の形態 2)

本発明の実施の形態 2 による音声認識装置の構成は実施の形態 1 と同様、図 1 の構成である。

【0057】

実施の形態 1 で説明したように不特定話者方式の音声認識装置で認識率や認識スピードを向上させることができたが、話者の『もしもし中村ですが、松下さんを、お願いします。』を『もし』『もし』『なか』『むら』『です』『が、』『

まつ』『した』『さん』『を、』『おね』『がい』『しま』『す。』のように全て分割し処理するには、装置の処理能力が必要とされ、小さな組み込み装置では処理スピードに影響を及ぼす事が考えられる。この問題を解決するため、名前の前後に来る言葉に規則性があることを利用し、これをあらかじめ登録しておき、前に来る言葉を開始信号、後に来る言葉を終了信号とする事により、よりトレーニングの精度をあげ処理スピードを向上させる。以下、その動作を説明する。

【0058】

実施の形態1と同様に、『もしもし中村ですが、松下さんを、お願いします。』を例にとって説明すると、図1において『もしもし中村ですが、松下さんを、お願いします。』なる音声信号がマイク1より信号処理部5に入力される。信号処理部5では、『もしもし中村ですが、松下さんを、お願いします。』の音声を或る任意の時間間隔で『もし』『もし』『なか』『むら』『です』『が、』『まつ』『した』『さん』『を、』『おね』『がい』『しま』『す。』と分割し、メモリ部16に出力する。また、その分割した音響データそれぞれに対して、音響データ識別子が信号処理部5より割り振られる。メモリ部16は、信号処理部5より入力された音響データ識別子に音響処理部7で生成された音響データを対応させ蓄積する。次に、メモリ部16は、この蓄積された音響データとそれに対応する音響データ識別子を単語認識部9に出力する。ここで、単語辞書格納部12に、あらかじめ名前の前後に発生しやすい言語、例えば『が』に代表される助詞や『さん』に代表される敬称を登録しておき、音素モデル格納部13から出力される音素データと共に言語モデル作成格納部14の中で作成し格納しておく。メモリ部16より『が』の音響データが単語識別部9に入力されたとき、単語識別部9は言語モデル作成格納部14に作成蓄積されている単語音響データと音響データとで識別動作を行い、或る任意の一致度以上の結果があった場合、信号処理部5に対して識別情報を出力する。信号処理部5は、開始信号として登録している単語識別子と認識信号の比較を行い、一致した場合、これを開始信号として記憶する。終了信号においても同様な動作を行う。これによりトレーニングとして使用する『まつした』の前後の文字『が』、『さん』が特定でき、信号処理部5は開始信号の後と終了信号の前までの音響データを言語モデル作成格納部14に

出力させる制御信号をメモリ部 16 に出力する。

【0059】

これによりメモリ部 16 より出力された『まつした』の音響データが言語モデル作成格納部 14 に蓄積され、実施の形態 1 と同様の効果が得られると共に、実施の形態 1 よりもトレーニングの精度をあげ処理スピードを向上させた音声認識装置を得ることができる。

【0060】

(実施の形態 3)

実施の形態 2 では、助詞をもとに開始信号を検知し、トレーニングを行ったが、様々な種類の助詞が存在し、登録にはメモリ容量が必要である。この問題を解決するため、特に日本語においてトレーニングする名前の前には必ず休止期間が存在するため、これを認識して開始信号とする事により、より精度の高いトレーニングを行う。構成および動作は、実施の形態 2 とまったく同様で、単語辞書格納部 12 に無音の単語データを登録し、言語モデル作成格納部 14 に無音の単語音響データを作成格納しておく。例として、『もしもし中村ですが、松下さんを、お願いします。』において、『もしもし』の次に無音区間が入っても、『もしもし』で開始信号、『中村ですが、』で開始信号、『松下さん』で終了信号、『を、』で開始信号、『お願いします。』で開始信号となり、信号のみに着目すると、開始信号→開始信号→終了信号→開始信号→開始信号となり、信号処理部 5 で、開始信号→開始信号、終了信号→開始信号を無視し、開始信号→終了信号の部分を検知すれば、トレーニングは可能となる。

【0061】

このように、トレーニングの精度を向上させ、かつ、単語辞書格納部 12 ならびに言語モデル作成格納部 14 のメモリ量を削減した音声認識装置を得ることができる。

【0062】

(実施の形態 4)

実施の形態 3 では、休止期間の検知を音声認識処理部 6 で行ったが、より処理能力の低い装置に対応するためには、機器のソフト処理を軽減する必要がある。

この問題を解決するため、検出部を信号処理部 5 に設けてハード処理で行うことにより、装置全体の負荷を軽減し、より高速な認識スピードを実現したものである。

【0063】

図 2、図 3 は、本発明の実施の形態 4 による音声認識装置を構成する信号処理部 5 の音声パス部を示すブロック図である。

【0064】

図 2、図 3 において、17 はフィルタ部、18 はゲインコントロール部、19 は A/D 変換部、20 は制御部、21 は電圧レベル検知回路である。

【0065】

このように構成された音声認識装置について、その動作を説明する。

【0066】

マイク 1 に入力された音声は、アナログ音声信号のままフィルタ部 17 に入力され、不要な信号成分を除去した後、ゲインコントロール部 18 に入力される。ゲインコントロール部 18 で任意のレベルに調整され、A/D 変換部 19 に入力される。A/D 変換部 19 でデジタル音声信号に変換され、次段の音響処理部 7 に入力される。本実施の形態では、図 3 の如く、フィルタ部 17 とゲインコントロール部 18 の間若しくはゲインコントロール部 18 と A/D 変換部 19 の間に、又は A/D 変換部 19 の後に電圧レベル検知回路 21 を設け、無音レベルを検知し、制御部 20 に検知信号を出力する。制御部 20 は、電圧レベル検知回路 21 から出力される検知信号を受け、メモリ部 16 に対して信号を出力する。以下は実施の形態 3 と同じ動作となる。

【0067】

このようにして、処理能力を軽減させ、認識スピードがより向上した音声認識装置を得ることができる。

【0068】

(実施の形態 5)

実施の形態 4 では、開始信号をハード処理により検知し、装置の処理を軽減させたが、ハード的な検知なので、周辺のノイズに対し誤った検知を行う恐れが生

じる。本実施の形態では、電圧レベル検知回路 21 のアナログ部では検知電圧の閾値を持たせ、デジタル部では任意の値を持たせ、閾値以上の電圧若しくは任意の値以上の値を検知した場合にのみ、制御部 20 に検知信号を出力する。

【0069】

これにより、ノイズ耐性を向上させた音声認識装置を得ることができる。

【0070】

(実施の形態 6)

実施の形態 1 から実施の形態 5 までは、話者の認識率の向上や認識スピード向上またはトレーニング精度の向上を図り、話者の利便性向上を図ったが、より小さな処理能力の装置に対して更なる認識スピードの高速化を図る必要がある。本実施の形態では、この問題を解決するため、単語辞書格納部 12 の格納の仕方を改善し、単語識別部 9 の識別速度を向上し、話者への利便性を向上する。構成、動作は図 1 の構成、動作と同じであり、単語辞書格納部 12 をもとに、その構成ならびに単語の読み出し方について説明を行う。

【0071】

図 4 は、単語辞書格納部 12 における単語データの一般的な例を示すデータ図である。各単語にはそれぞれ話者が登録した名前が格納されており、認識動作が行われるに当たり、上から順次に全て言語モデル作成格納部 14 に出力されて行く。

【0072】

図 5 は、本発明の実施の形態 6 における単語データの配置を示すデータ図である。図 5 においては、単語の先頭部分とその他の部分とに分けて格納し、先頭部分の文字が同じ物をグループ化する。以下、一連の動作を図 1 を使い説明すると、話者がマイク 1 に向かって一例として『まつした』と発音した場合、様々な処理を経て単語識別部 9 に入力される。これを受け、単語辞書格納部 12 から音響データが順次出力されて行くが、まず、先頭文字のみを出力し、言語モデル作成格納部 14 に入力される。言語モデル作成格納部 14 では、音素モデル格納部 13 から出力される音素データをもとに先頭文字のみの単語音響データを作成し、単語識別部 9 へ出力する。言語モデル作成格納部 14 では、音響データが 1 文字

であるため、短時間で単語音響データを作成する事が可能である。単語識別部 9 は、音響処理部 7 からの音響データを識別し、単語識別子を識別情報として出力する。これを受けた信号処理部 5 は、単語辞書格納部 12 に対し、識別情報から判別されるグループ番号を出力し、単語辞書格納部 12 は、言語モデル作成格納部 14 に対し、特定のグループ番号のみの単語データを出力する。

【0073】

以上のように、単語辞書格納部 12 に登録されている特定のグループのみ音響データ化されるため、認識スピードが向上し、名前の格納の仕方により単語辞書格納部 12 内のメモリ容量を削減した音声認識装置を得ることができる。

【0074】

(実施の形態 7)

実施の形態 6 では、単語辞書格納部 12 から先頭文字を読み出して識別を行ったが、更なる高速化を図るため、あらかじめ単語辞書格納部 12 内の先頭文字と音素モデルとから先頭文字の単語音響データを作成し、言語モデル作成格納部 14 に格納しておき、識別動作時、単語辞書格納部 12 からの単語データの呼び出しや音素モデル格納部 13 からの音素データの呼出し、また、それらによる単語音響データの作成時間を省くことにより、更なる高速化が図れる。

【0075】

(実施の形態 8)

実施の形態 6 では、先頭文字のみ切り分けて単語辞書格納部 12 に格納する構成としたが、単語辞書格納部 12 の名前の登録内容は苗字だけではなく、名前も合わせて格納され、メモリ量が増大する恐れがある。本実施の形態は、この問題を解決するものであり、これを図 6 を用いて説明する。図 6 は苗字の先頭文字とその他の部分と名前とを分けて格納する場合を示すデータ図である。

【0076】

図 6 に示すように、苗字の先頭文字とその他の部分と名前とを分けて格納するようにしたことにより、メモリ量を更に削減した音声認識装置を得ることができる。

【0077】

(実施の形態 9)

実施の形態 1 における単語辞書格納部 12 から音響データを呼び出す方法は、単語辞書格納部 12 のアドレスの高位から低位へ、若しくは低位から高位と全てに渡って単純に読み出しを行い、過去一度も使用されていないものまでも言語モデル化し識別しており、この場合は高い処理能力と時間が要求される。この問題を解決するため、単語識別部 9 の識別動作で生成出力される識別情報に含まれる一致度情報を利用し、信号処理部 5 で、一致度が一番高い単語識別子を持つ単語データのみに頻度 1 を与え使用のたびごとに加算して行き、信号処理部内 5 に格納、蓄積する。この蓄積した頻度情報をもとに、単語辞書格納部 12 のメモリ（図示せず）に格納されている単語データを頻度順に並べ替え、次に識別動作を行う際、頻度の高い順に言語モデル作成格納部 14 へ出力され、単語音響データに変換された後、単語識別部 9 で識別動作を行う。単語識別部 9 は識別情報を出力するが、信号処理部内 5 は入力された識別情報内の一致度を監視しつつ、或る任意の一致度以下になった場合は、それまで識別情報として蓄積した単語識別子にしたがった単語を表示装置 4 に出力させる。

【0078】

このようにして使用頻度の高い単語から識別され、また、表示する単語データの頻度に閾値を設ける事により、より高速な認識動作を行える音声認識装置を得ることができる。

【0079】**(実施の形態 10)**

実施の形態 9 では、表示させる単語の選別を一致度で行ったが、本実施の形態では、使用頻度そのものに閾値を設け、或る任意の値以下の単語データは言語モデル作成格納部 14 に出力しない事により、認識動作の高速化が図れる音声認識装置を得ることができる。

【0080】**(実施の形態 11)**

実施の形態 9 や実施の形態 10 では、装置の使用頻度が高ければ良いが、低い場合には、登録したが表示されないという問題が発生する可能性がある。この間

題を解決するため、使用頻度の高い順に任意の単語数ごとにブロック化し、頻度の高いブロックから音響データを出力させ、ブロック毎に表示させる事で、低い頻度の音声データが入力されても確実に表示できる音声認識装置を得ることができる。これを図7に示す。図7は、単語辞書格納部12の単語データ配列を使用頻度の高い順に示すデータ図である。

【0081】

(実施の形態12)

実施の形態9、実施の形態10や実施の形態11では、過去に使用頻度が高いが現在はほとんど使用しない単語データがある場合、話者が目的とする単語をすぐに表示できないという問題がある。この問題を解決するため、信号処理部5に時計機能を有し、或る任意の時間が経過した頻度が高い単語データを頻度を下げ配置しなおすことにより、より高速な利便性に優れた音声認識装置を得ることができる。

【0082】

(実施の形態13)

不特定話者方式、特定方式を問わず、音声認識装置全般において、ある特定の言語に対して何度も誤認識を繰り返す傾向がある。この問題を解決するため、本実施の形態は、信号処理部5のメモリを使用することにより、一度誤った単語は二度と表示させないようにしたものである。以下、この動作の説明を行うが、本実施の形態による音声認識装置の構成は図1と同じ構成である。

【0083】

図1にて、マイク1に音声が入力され、マイク1からのアナログ音声信号が信号処理部5に入力される。このアナログ音声信号は、信号処理部5では、最終的にA/D変換され、音響処理部7へデジタル音声信号として出力される。一方、この音声信号を信号処理部5のメモリに蓄積する。その後の動作として、実施の形態1で説明した一連の動作が行われ、単語識別部9は、単語識別子を含む識別情報を信号処理部5に出力する。信号処理部5は、単語識別子を含む識別情報を先にメモリに蓄積した音声信号と関連づけて蓄積する。この識別情報にもとづき表示装置4に単語データを表示させるが、表示装置4に話者の意図しない単語が

表示された場合、話者は入力装置 3 にて表示を消す操作を必ず行う。この操作により信号処理部 5 はメモリに蓄積した識別情報と単語識別子が誤ったものであると認識し、誤ったことを示す情報も、先に蓄積した音声信号と単語識別子と関連づけて蓄積する。次に、別の機会に話者が前回と同じ発音をした場合、音声信号は前述と同様に A/D 変換され、信号処理部 5 のメモリに蓄積される。信号処理部 5 はこのとき、以前蓄積した音声信号と同じであるか否かを判別する。同時に音声信号は、音響処理部 7 へ出力され、その後一連の処理の後、結果的に単語識別部 9 より単語識別子を含む識別情報が出力される。信号処理部 5 は、この単語識別子を認識し、前回蓄積した単語識別子と同一であるならば、再び誤認識をしたと判定し、表示装置 4 に対して単語識別子に対応する単語データの表示は出力せず、次に来る識別情報に含まれる単語識別子に基づいた単語データを表示させる。

【0084】

このようにして話者が一度誤認識したと判定した単語データは二度と表示させないという利便性に優れた音声認識装置を得ることができる。

【0085】

(実施の形態 14)

実施の形態 13 では、信号処理部 5 のメモリを使用したか、信号処理部 5 は、表示装置 4 への表示や入力装置 3 の監視の他、様々な制御でメモリを使用しているため、信号処理部 5 のメモリだけでは容量が足らなくなる恐れがある。本実施の形態は、この問題を解決するため、音響処理部 7 に繋がるメモリ部 16 を使用することにより、実施の形態 13 と同じ効果を得るものである。以下その動作の説明を行うが、本実施の形態による音声認識装置の構成は図 1 の構成である。

【0086】

マイク 1 に音声が入力され、マイク 1 からのアナログ音声信号は信号処理部 5 に入力される。このアナログ音声信号は、信号処理部 5 では最終的に A/D 変換され、音響処理部 7 へデジタル音声信号として出力される。音響処理部 7 で特徴量を抽出し、特徴量はメモリ部 16 と単語識別部 9 に出力され、メモリ部 16 はこの特徴量を蓄積する。その後の動作として、実施の形態 1 で説明した一連の動

作が行われ、単語識別部 9 は、単語識別子を含む識別情報を信号処理部 5 に出力する。信号処理部 5 は、この識別情報にもとづき、表示装置 4 に単語データを表示させるが、表示装置 4 に話者の意図しない単語が表示された場合、話者は入力装置 3 にて表示を消す操作を必ず行う。この操作により、信号処理部 5 は、メモリ部 16 に蓄積した音響データと単語識別子が誤ったものであると認識し、それを記憶する。次に、別の機会に話者が前回と同じ発音をした場合、音声信号は前述と同様に A/D 変換され、音響処理部 7 へ出力され、メモリ部 16 に蓄積される。信号処理部 5 は、前回蓄積した音響データと今回蓄積した音響データとが同一であるか否かを判別する。この場合同じ言葉が発せられたので、信号処理部 5 は同一と判定する。その後の一連の処理の後、結果的に単語識別部 9 より単語識別子を含む識別情報が出力される。信号処理部 5 は、この単語識別子を認識し、前回蓄積した単語識別子と同一であるならば、再び誤認識をしたと判定し、表示装置 4 に対して、単語識別子に対応する単語データの表示は出力せず、次に来る識別情報に含まれる単語識別子に基づいた単語データを表示させる。

【0087】

このようにして実施の形態 13 と同様な効果が得られるが、信号処理部 5 の負荷を減らし、また、特徴量抽出後のデータであるため、より少ないメモリ量で実現できる音声認識装置を得ることができる。

【0088】

(実施の形態 15)

音声認識技術を使った装置が各国に広まる中、製造メーカは製造コスト削減のため、仕向け地全ての音素モデルを機器に搭載し、ユーザのキー操作で言語にあった音素モデルを選択する必要がある。しかしながら、今後の音声認識技術と音声合成技術のレベルが進むと、全くキーを持たない装置（つまり入力装置の無い装置）が出てくることが予想され、音素モデルを製造メーカ側で仕向け地に従い搭載しなければならなくなり、製造コストのアップに繋がる。本実施の形態は、この問題を解決するため、仕向け地ごとのある特定の言葉に対して、その言葉を単語辞書格納部 12 にあらかじめ持たせることによって、また、音素モデル格納部 13 を信号処理部から制御する事によって、ユーザが機器を使用する最初にそ

の言葉を発音する事により自動的な言語選択を可能にするのもである。以下、その動作について図 8 を参照しながら説明を行う。

【0089】

図 8 は、本発明の実施の形態 15 による音声認識装置を示すブロック図である。図 8 の構成が図 1 と異なるところは、図 8 では図 1 の入力装置 3 が無い点である。

【0090】

音声認識装置が製品として出荷され話者が使用する前は、通常、単辞書格納部 12 には、何もデータがない。また、音素モデルには、それぞれ各国の音素データが格納されている。本実施の形態では、単語辞書格納部 12 に、あらかじめ各国の任意の同じ意味の言葉、例えば日本語では『いち』、英語では『One』、ドイツでは『Eine』等を製品出荷時に格納しておく。話者（ユーザ）は、製品を入手した後、日本語の数字の『いち』に相当する言葉をマイク 1 より各国の言葉で入力する事によって、これまで説明してきたと同様な動作を繰り返し、単語識別部 9 よりどの国の言語かの識別情報が出力され、信号処理部 5 に入力される。信号処理部 5 は、音素モデル格納部 13 に対し、制御信号を出力し、音素モデル格納部 13 は、該当する言語に対応する音素モデルが格納されている部分以外のゲートを閉じ、該当する言語に対応する音素モデルのみを出力して行く。また、他の言語に変更する時は、選択された言語で特定の言葉を入力すると、一連の動作により、信号処理部 5 から制御信号が出力され、音素モデル格納部 13 の全ての言語のゲートが開き、変更が行える。

【0091】

以上により、入力装置を持たない装置においても言語選択が可能な音声認識装置を得ることができる。

【0092】

【発明の効果】

以上説明したように本発明の請求項 1 に記載の音声認識装置によれば、キーボードなどの入力装置と、認識結果を表示する表示装置と、入力装置と表示装置が接続され、音声認識装置全体の制御を行う信号処理部と、信号処理部より入力さ

れた音声信号を分割した音響データと内部で保持している複数の単語音響データとを比較し、分割した音響データに一致する単語音響データの単語識別子を含む識別情報音声認識の結果として信号処理部へ出力する音声認識処理部とを有する不特定話者方式の音声認識装置であって、音声認識処理部は、既に登録されている音素モデルから生成される単語音響データと話者が発した名前から生成される音響データとを逐次比較し、単語音響データに合致する音響データが生成されたとき、生成された音響データに対応する音響データ識別子をトレーニング信号として蓄積することにより、話者に意識させずにトレーニングを行うことができるので、話者の負担になるトレーニングを排除することができ、また、不特定話者方式でありながら、特定話者方式の特徴である高い認識率と認識スピードを得ることができるという有利な効果が得られる。

【0093】

請求項2に記載の音声認識装置によれば、請求項1に記載の音声認識装置において、信号処理部は、名前が発音される前後の言葉を記憶するメモリ部を有し、名前の前の言葉を開始信号、後の言葉を終了信号とすることにより、名前が発音される前後の言葉の一致度が向上し、トレーニング精度を更に向上させることができるという有利な効果が得られる。

【0094】

請求項3に記載の音声認識装置によれば、請求項2に記載の音声認識装置において、信号処理部は、日本語において名前の前に必ず存在する無音を記憶するメモリ部を有し、無音を検知して開始信号とすることにより、無音を名前の開始信号とすることができるので、メモリ部における必要メモリ量を低減することができると共に、名前のトレーニング精度を更に向上させることができるという有利な効果が得られる。

【0095】

請求項4に記載の音声認識装置によれば、請求項3に記載の音声認識装置において、信号処理部は、無音を検知する検知部と、検知した無音を開始信号とする制御部とを有することにより、無音検知を信号処理部で行うので、音声認識処理部の負荷を軽減することができ、音声認識処理部におけるトレーニング処理をよ

り向上させることができるという有利な効果が得られる。

【0096】

請求項5に記載の音声認識装置によれば、請求項4に記載の音声認識装置において、信号処理部は、無音検知のためのスレッシュホールドレベルを検知部に設けたことにより、雑音による誤動作発生を防ぐことができるという有利な効果が得られる。

【0097】

請求項6に記載の音声認識装置によれば、請求項1に記載の音声認識装置において、音声認識処理部は、単語を単語辞書格納部に格納する際、単語の先頭部分とその他の部分に分け、先頭部分が同じ物をグループ化して格納することにより、先頭部分が同じグループのみを言語モデル化して識別を行うようにすることができるので、識別動作の高速化を図ることができ、単語辞書格納部のメモリ量の削減を図ることができるという有利な効果が得られる。

【0098】

請求項7に記載の音声認識装置によれば、請求項6に記載の音声認識装置において、音声認識処理部は、あらかじめ単語辞書格納部内の先頭文字と音素モデルから先頭文字の単語音響データを作成し、言語モデル作成格納部に格納することにより、単語辞書格納部からの単語データの呼び出しや言語モデル作成格納部からの音素データの呼び出し、またそれらによる単語音響データの作成時間を省くことができ、更なる高速化が図れるという有利な効果が得られる。

【0099】

請求項8に記載の音声認識装置によれば、請求項6に記載の音声認識装置において、音声認識処理部は、単語辞書の構成を先頭文字と姓と名前とにブロック分けすることにより、ブロック分けが更に進められ、使用メモリ量をさらに削減することができるという有利な効果が得られる。

【0100】

請求項9に記載の音声認識装置によれば、請求項1に記載の音声認識装置において、音声認識処理部は、言語モデル作成格納部に格納された各単語音響データに対して、使用されたとき単語識別部より出力される一致度で一番高いものに頻

度1を与え、使用のたびに加算して重み付けを行うことにより、重み付けの高い順に単語音響データを並べ替えることができるので、音声認識時の高速化を図ることができるという有利な効果が得られる。

【0101】

請求項10に記載の音声認識装置によれば、請求項9に記載の音声認識装置において、音声認識処理部は、或る任意の頻度以上の単語音響データのみを使って認識動作を行わせることにより、或る任意の値以下の頻度を持つ単語音響データは言語モデル作成格納部に出力しないので、音声認識における高速化を図ることができるという有利な効果が得られる。

【0102】

請求項11に記載の音声認識装置によれば、請求項9に記載の音声認識装置において、音声認識処理部は、使用頻度の高い順に任意の単語数ごとにブロック化し、使用頻度の高いブロックの単語音響データを出力させ、ブロック毎に表示させることにより、使用頻度が低い音声が入力されても認識が可能で、またブロック毎の識別動作で表示する事により高速な認識動作を行うことができるという有利な効果が得られる。

【0103】

請求項12に記載の音声認識装置によれば、請求項9乃至11のいずれか1に記載の音声認識装置において、信号処理部は時計機能を有し、音声認識処理部は信号処理部から通知される時刻に基づいて使用頻度を算出する期限を設けたことにより、使用頻度が最も高いものであっても期限外ならば低位に位置することができるので、ユーザに取って近似点で最も使用頻度が高いものから順に認識動作を行うようにすることができ、より高速化を図ることができるという有利な効果が得られる。

【0104】

請求項13に記載の音声認識装置によれば、請求項1に記載の音声認識装置において、信号処理部は、認識動作後に表示装置に表示された結果がユーザの意図する結果と異なった場合には異なったことを示す情報を内蔵するメモリ部に蓄積し、次ぎに同じ発音があった場合に、異なったことを示す情報をもとに、一度間

違った情報は二度と表示装置に表示しないことにより、一度間違った情報は二度と表示装置に表示しないので、ユーザにとっての利便性が向上するという有利な効果が得られる。

【0105】

請求項14に記載の音声認識装置によれば、請求項1に記載の音声認識装置において、信号処理部は、認識動作後に表示装置に表示された結果がユーザの意図する結果と異なった場合には異なったことを示す情報を音声認識処理部のメモリ部に蓄積し、次ぎに同じ発音があった場合に、異なったことを示す情報をもとに、一度間違った情報は二度と表示装置に表示しないことにより、請求項13と同様の有利な効果が得られると共に、音声認識処理部のメモリ部に情報を蓄積するので、信号処理部の負荷およびメモリ量を減らすことができるという有利な効果が得られる。

【0106】

請求項15に記載の音声認識装置によれば、請求項1に記載の音声認識装置において、音声認識処理部は、仕向け地ごとの或る特定の言葉を単語辞書格納部にあらかじめ持たせることにより、ユーザが装置を使用する最初に或る特定の言葉を発音して、自動的に言語を選択することができるという有利な効果が得られる。

【図面の簡単な説明】

【図1】

本発明の実施の形態1による音声認識装置を示すブロック図

【図2】

本発明の実施の形態4による音声認識装置を構成する信号処理部の音声パス部を示すブロック図

【図3】

本発明の実施の形態4による音声認識装置を構成する信号処理部の音声パス部を示すブロック図

【図4】

単語辞書格納部における単語データの一般的な例を示すデータ図

【図 5】

本発明の実施の形態 6 における単語データの配置を示すデータ図

【図 6】

苗字の先頭文字とその他の部分と名前とを分けて格納する場合を示すデータ図

【図 7】

単語辞書格納部の単語データ配列を使用頻度の高い順に示すデータ図

【図 8】

本発明の実施の形態 15 による音声認識装置を示すブロック図

【図 9】

特定話者方式を用いる従来の音声認識装置を示すブロック図

【図 10】

図 9 の音声認識処理部を示すブロック図

【図 11】

図 10 の単語音響データ格納部を示すブロック図

【図 12】

不特定話者方式を用いる従来の音声認識装置を示すブロック図

【図 13】

図 12 の音声認識処理部を示すブロック図

【図 14】

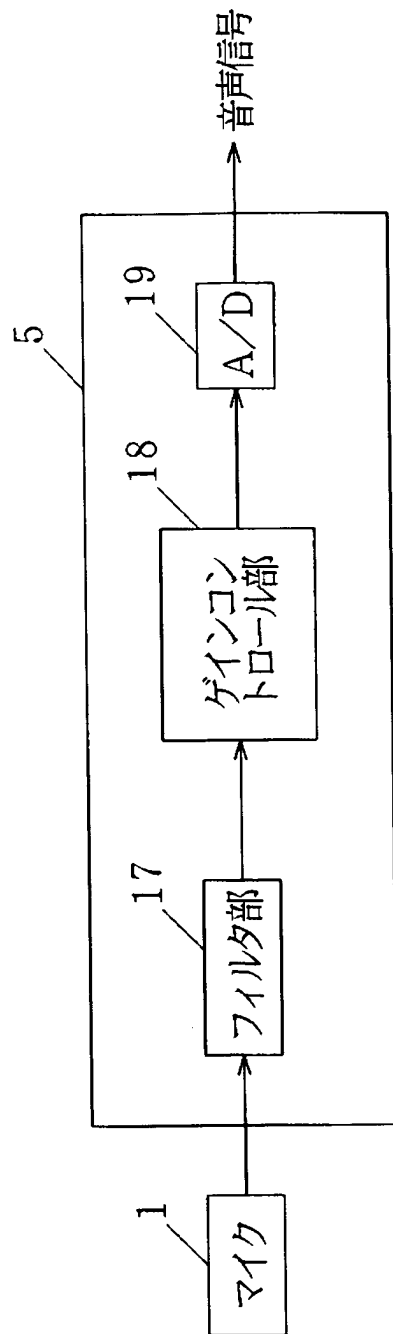
図 13 の単語辞書格納部を示すブロック図

【符号の説明】

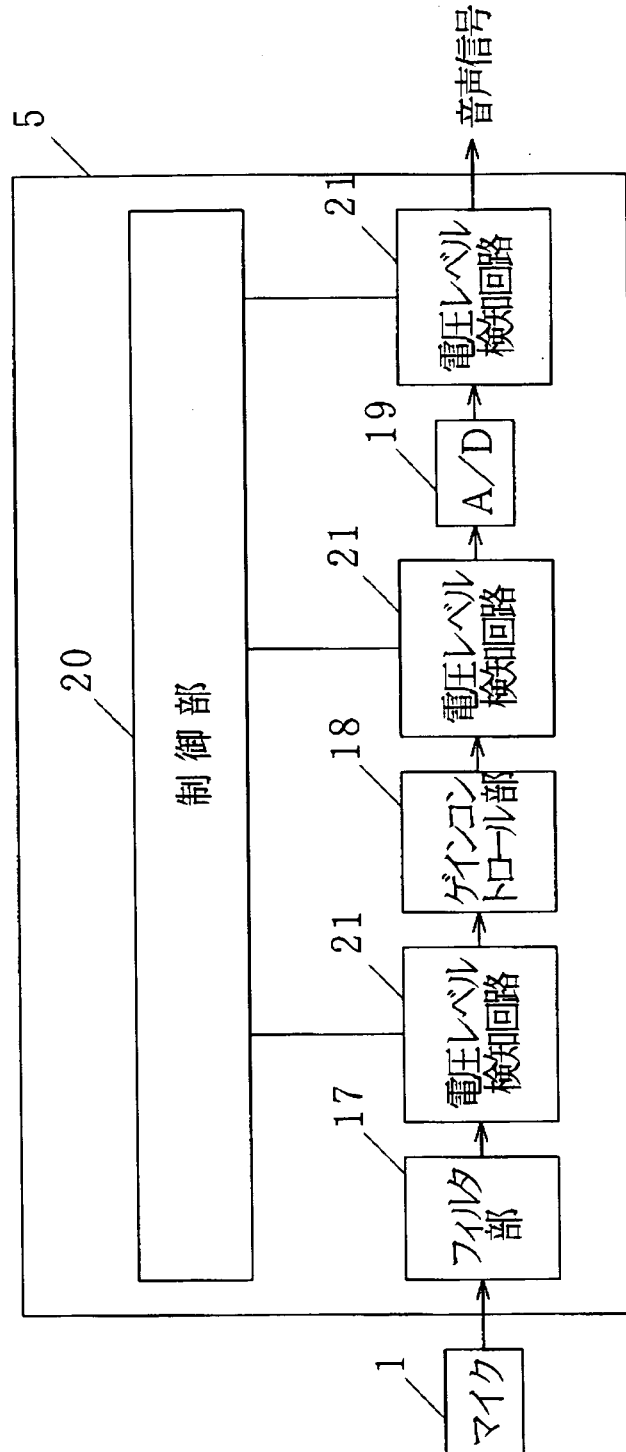
- 1 マイク
- 2 スピーカ
- 3 入力装置
- 4 表示装置
- 5 信号処理部
- 6 音声認識処理部
- 7 音響処理部
- 9 単語識別部

- 1 2 単語辞書格納部
- 1 3 音素モデル格納部
- 1 4 言語モデル作成格納部
- 1 6 メモリ部
- 1 7 フィルタ部
- 1 8 ゲインコントロール部
- 1 9 A／D変換部
- 2 0 制御部
- 2 1 電圧レベル検知回路

【図 2】



【図 3】



【図 4】

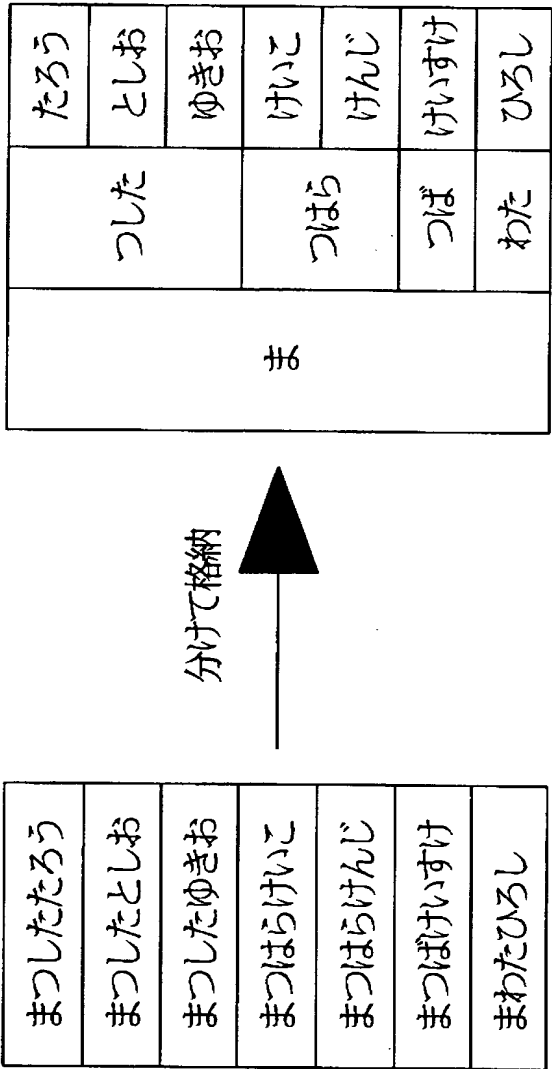
上から順次 全ての単語が言語
モデル作成格納部へ出力される

	名前
単語識別子	まつした
単語識別子	まつば
単語識別子	まわた
単語識別子	みうら
単語識別子	みさか
・ ・ ・	・ ・ ・
単語識別子	わたなべ
単語識別子	わたべ

【図 5】

	グループ	先頭文字	その他の文字
単語識別子	グループ1	ま	つした
単語識別子			つば
単語識別子			わた
単語識別子	グループ2	み	うら
単語識別子			さか
・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・
単語識別子	グループN	わ	たなべ
単語識別子			たべ

【図 6】

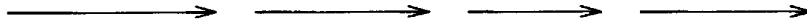


【図 7】

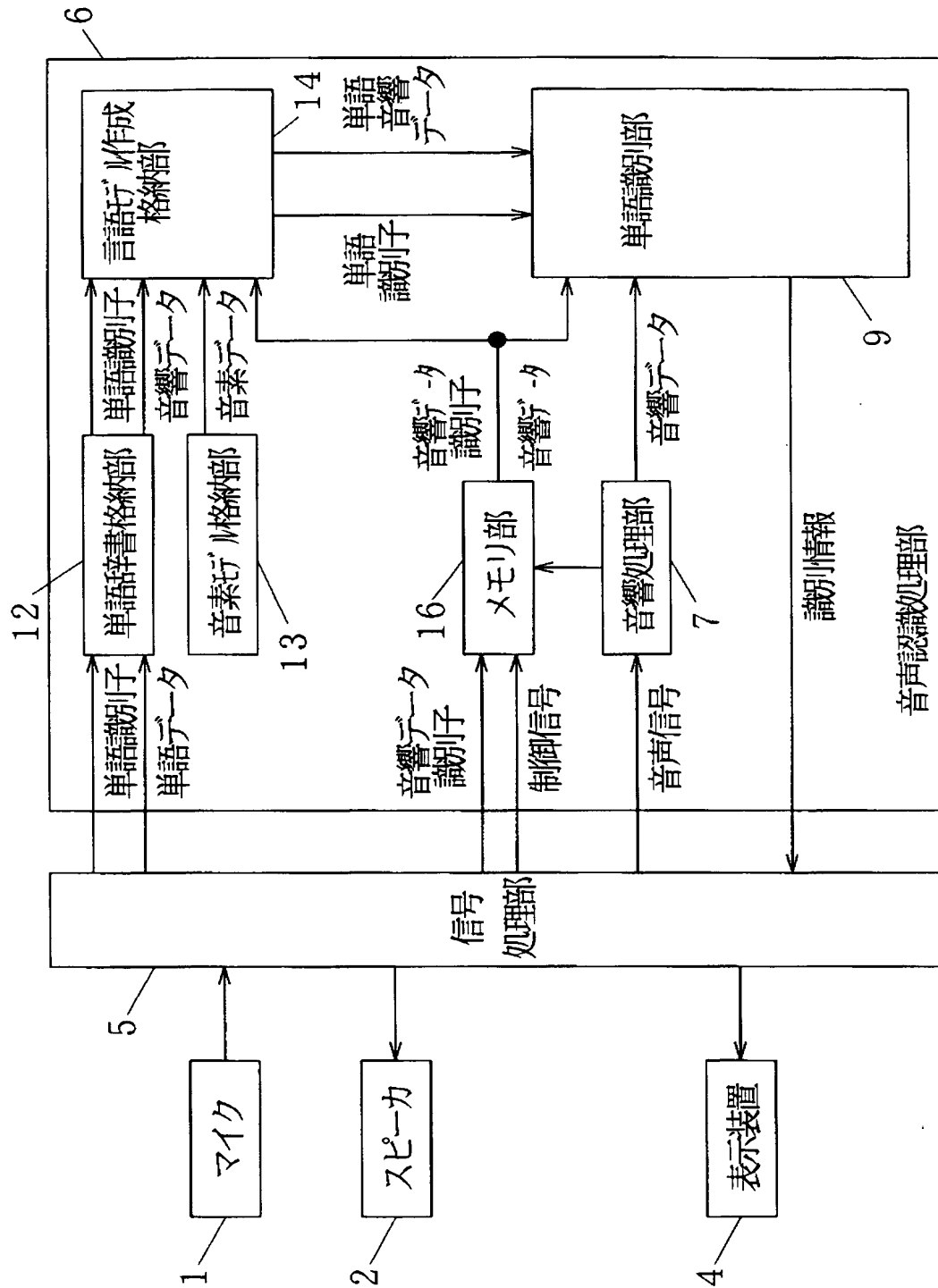
	グループ	使用頻度	単語
単語識別子	グループ 1	20	単語データ
単語識別子		18	単語データ
単語識別子		18	単語データ
単語識別子	グループ 2	16	単語データ
単語識別子		15	単語データ
単語識別子		14	単語データ
・ ・ ・	・ ・ ・		・ ・ ・
単語識別子	グループ N	0	単語データ
単語識別子		0	単語データ
単語識別子		0	単語データ

グループ 1 が最初に音響データとして出力され識別され表示される

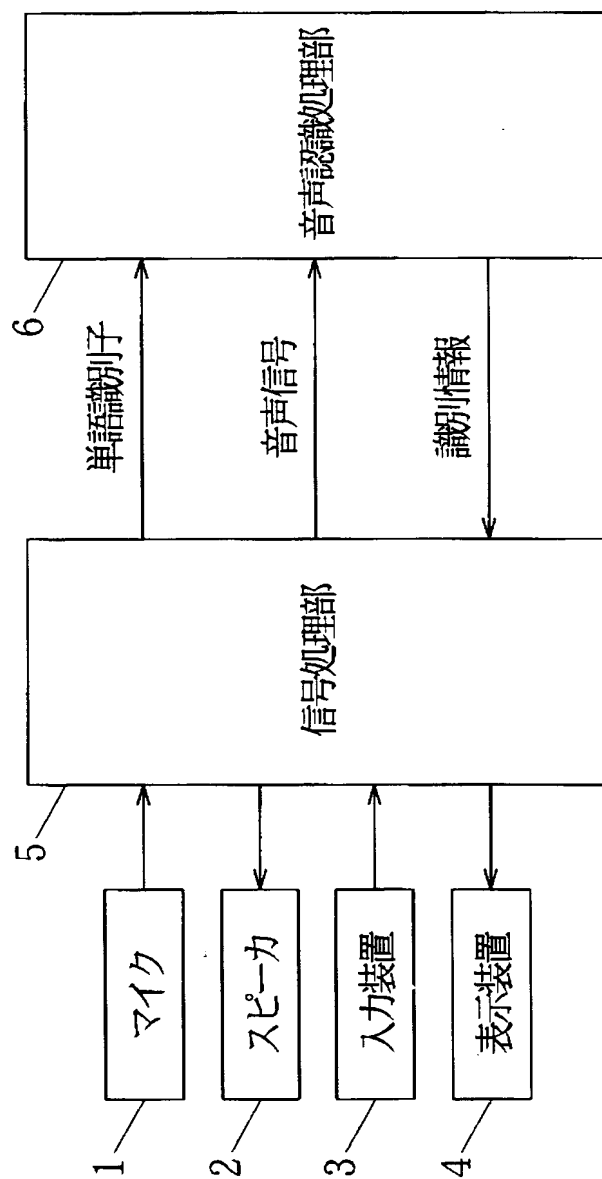
グループ 1 で該当がなかった場合、グループ 2 が次に音響データとして出力され識別され表示される



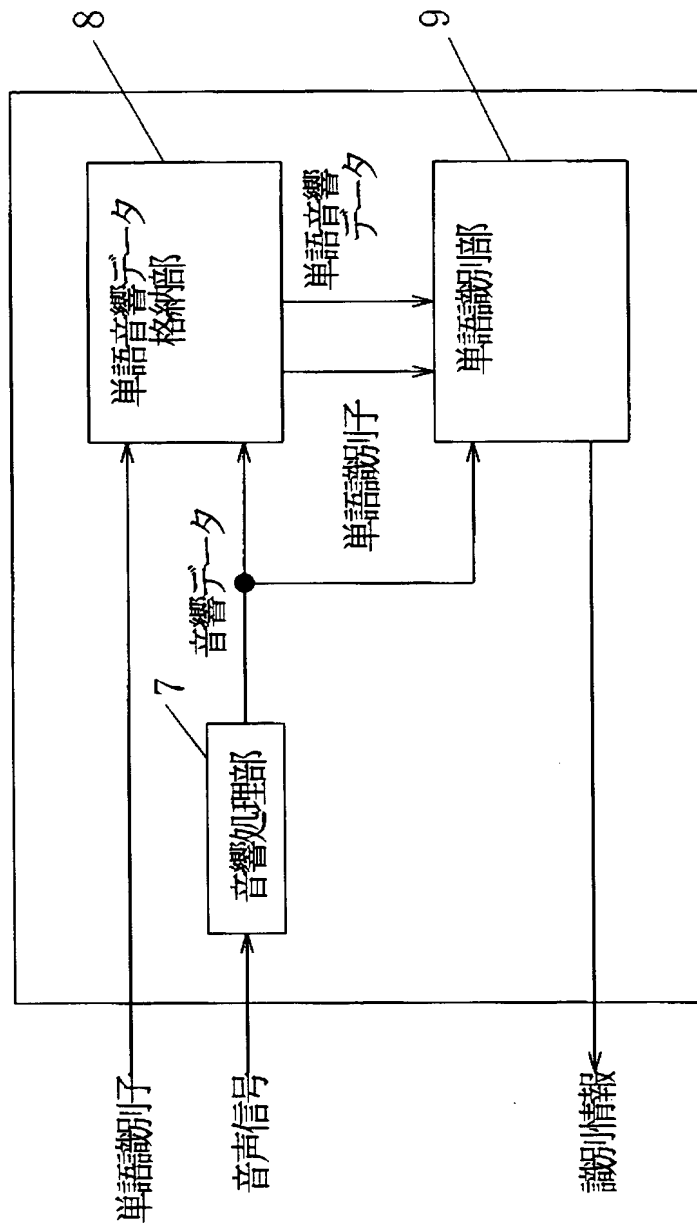
【図 8】



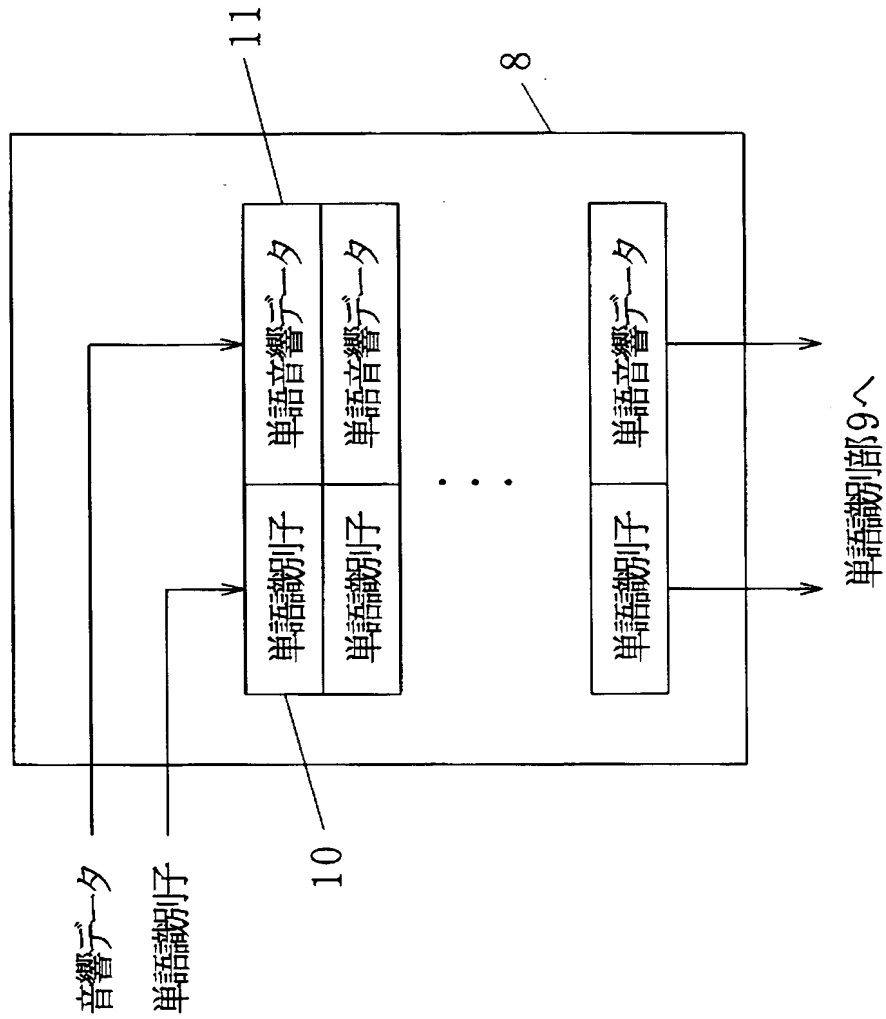
【図 9】



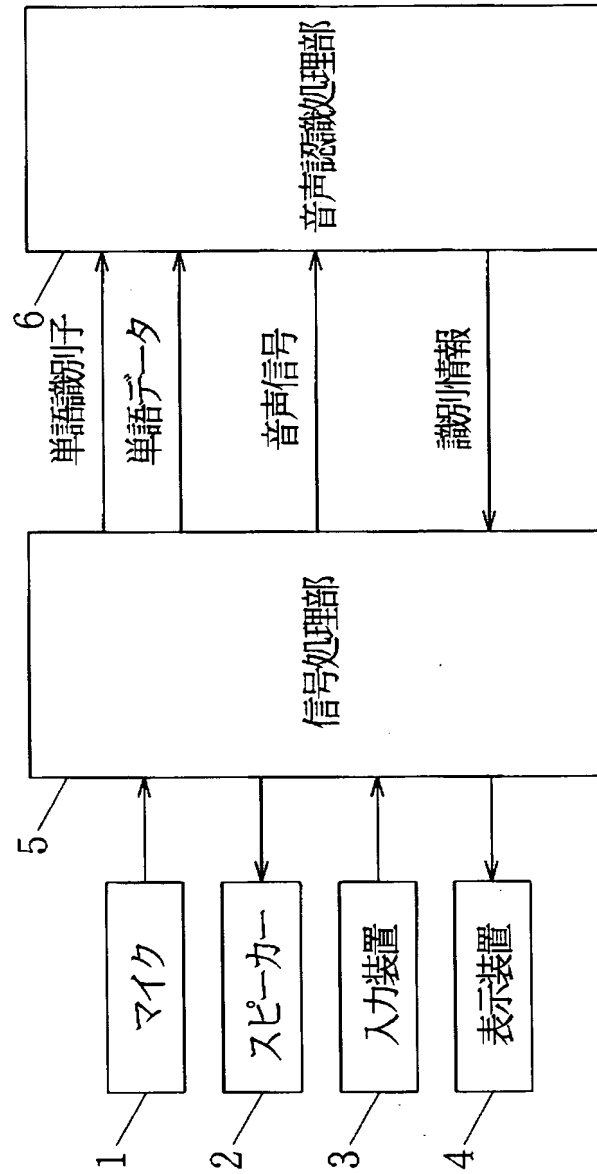
【図 10】



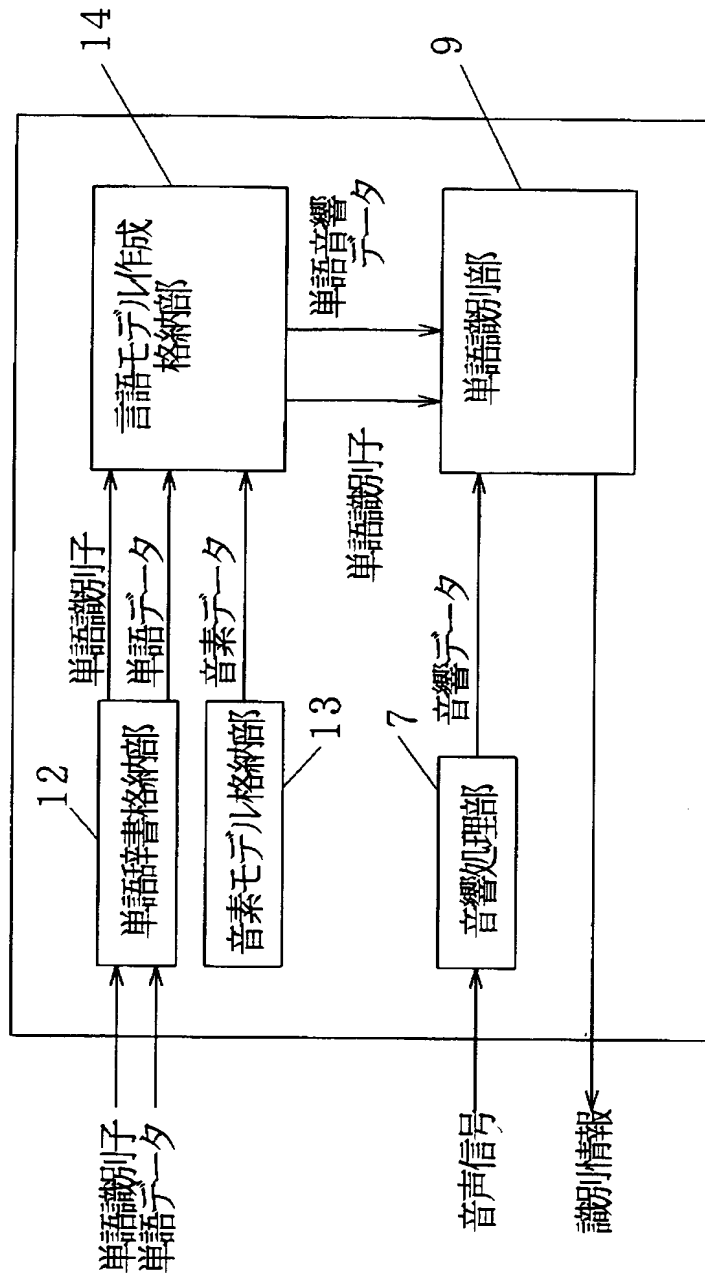
【図 1 1】



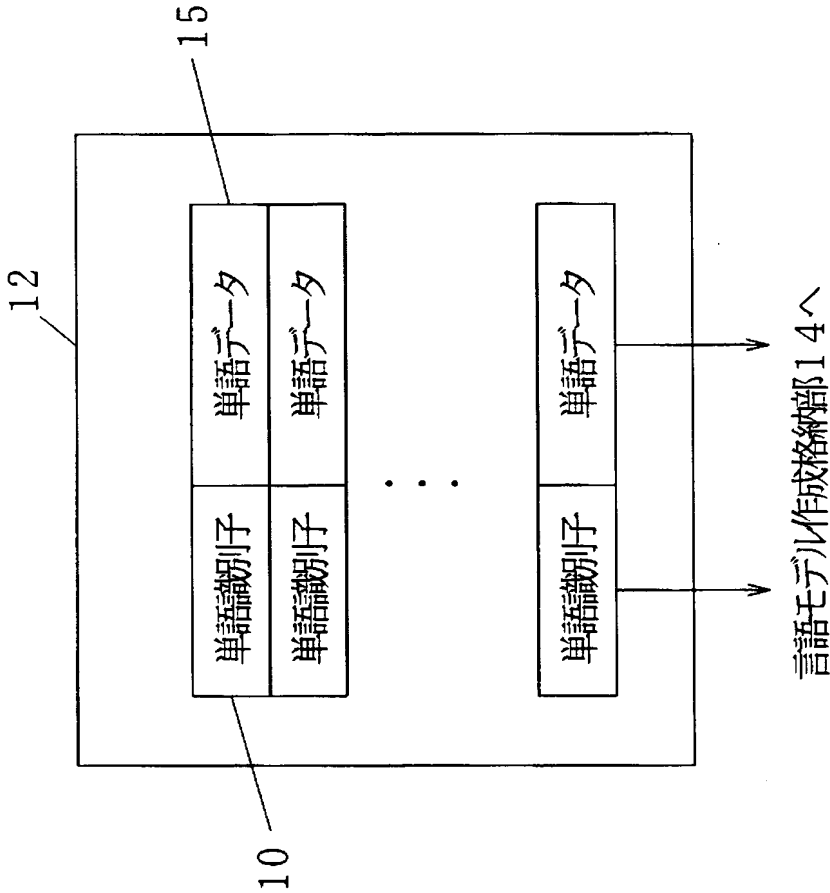
【図 12】



【図 13】



【図 1 4】



【書類名】 要約書

【要約】

【課題】 話者に意識させることなくトレーニングを行うことができ、使用して行くにつれて特定話者方式の認識率と認識スピードを高めることができる音声認識装置を提供することを目的とする。

【解決手段】 入力装置 3 と、表示装置 4 と、入力装置 3 と表示装置 4 が接続され、全体制御を行う信号処理部 5 と、信号処理部 5 からの音声信号を分割した音響データと複数の単語音響データとを比較し、分割音響データに一致する単語音響データの識別情報を信号処理部 5 へ出力する音声認識処理部 6 とを有する不特定話者方式の音声認識装置であって、音声認識処理部 6 は、音素モデルから生成される単語音響データと話者が発した名前から生成される音響データとを逐次比較し、単語音響データに合致する音響データが生成されたとき、生成された音響データに対応する音響データ識別子をトレーニング信号として蓄積する。

【選択図】 図 1

特願 2 0 0 2 - 2 0 8 3 8 6

出 願 人 履 歷 情 報

識別番号

[0 0 0 0 0 5 8 2 1]

1. 変更年月日

1 9 9 0 年 8 月 2 8 日

[変更理由]

新規登録

住 所

大阪府門真市大字門真 1 0 0 6 番地

氏 名

松下電器産業株式会社